Working Papers of the Priority Programme 1859
Experience and Expectation.
Historical Foundations of Economic Behaviour
Edited by Alexander Nützenadel und Jochen Streb

No 9 (2019, March)

Christoph Behrens

Evaluating the Joint Efficiency of German Trade Forecasts. A nonparametric multivariate approach

Arbeitspapiere des Schwerpunktprogramms 1859 der Deutschen Forschungsgemeinschaft
„Erfahrung und Erwartung. Historische Grundlagen ökonomischen Handelns“ /
Working Papers of the German Research Foundation's Priority Programme 1859
"Experience and Expectation. Historical Foundations of Economic Behaviour"

Published in co-operation with the documentation and publication service of the Humboldt University, Berlin (https://edoc.hu-berlin.de).

ISSN: 2510-053X

Redaktion: Alexander Nützenadel, Jochen Streb, Sebastian Schöttler

V.i.S.d.P.: Alexander Nützenadel, Jochen Streb

SPP 1859 "Erfahrung und Erwartung. Historische Grundlagen ökonomischen Handelns"
Sitz der Geschäftsführung:
Humboldt-Universität
Friedrichstr. 191-193, 10117 Berlin
Tel: 0049-30-2093-70615, Fax: 0049-30-2093-70644
Web: https://www.experience-expectation.de
Koordinatoren: Alexander Nützenadel, Jochen Streb
Assistent der Koordinatoren: Sebastian Schöttler

Recommended citation:

Behrens, Christoph (2019*): Evaluating the Joint Efficiency of German Trade Forecasts. A nonparametric multivariate approach.* Working Papers of the Priority Programme 1859 "Experience and Expectation. Historical Foundations of Economic Behaviour" No 9 (March), Berlin

Evaluating the Joint Efficiency of German Trade Forecasts
-
A nonparametric multivariate approach

Christoph Behrens[a]

March 2019

Abstract

I analyze the joint efficiency of export and import forecasts by leading economic research institutes for the years 1970 to 2017 for Germany in a multivariate setting. To this end, I compute, in a first step, multivariate random forests in order to model links between forecast errors and a forecaster's information set, consisting of several trade and other macroeconomic predictor variables. I use the Mahalanobis distance as performance criterion and, in a second step, permutation tests to check whether the Mahalanobis distance between the predicted forecast errors for the trade forecasts and actual forecast errors is significantly smaller than under the null hypothesis of forecast efficiency. I find evidence for joint forecast inefficiency for two forecasters, however, for one forecaster I cannot reject joint forecast efficiency. For the other forecasters, joint forecast efficiency depends on the examined forecast horizon. I find evidence that real macroeconomic variables as opposed to trade variables are inefficiently included in the analyzed trade forecasts. Finally, I compile a joint efficiency ranking of the forecasters.

JEL classification: C53; F17; F47

Keywords: Trade forecasts; German economic research institutes; Forecast efficiency; Multivariate random forests

Address:

[a] Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany. E-mail: c.behrens@hsu-hh.de

Acknowledgement:
I thank Christian Pierdzioch for very helpful and constructive feedback. I thank the participants of the "Futures Past. Economic Forecasting in the 20th and 21st Century" conference in Hamburg and of the "13th ifo Dresden Workshop on Macroeconomics and Business Cycle Research" in Dresden for comments and remarks. I thank the German Research Foundation for financial support (Exploring the experience expectation nexus in macroeconomic forecasting using computational text analysis and machine learning; Project number: 275693836).

1 Introduction

When forming macroeconomic expectations, economic agents tend to rely on professional forecasts (Carroll, 2003). The optimality of these forecasts is therefore of crucial importance when it comes to policy making or investment decisions. Generally, a forecast is defined as optimal if it is unbiased and efficient, i.e., if its forecast error is unpredictable (Mincer and Zarnowitz, 1969). Correctly predicting trade developments plays a decisive role, as these dynamics greatly influence output growth and price levels through inflationary pressures from import prices (D'Agostino et al., 2017). Furthermore, many macroeconomic forecasters use a disaggregated approach when forming an economic growth (GDP) forecast. In other words, research institutes form individual forecasts for all components of the GDP, such as exports or domestic demand, and conflate the results to a prediction of overall output.[1] Hence, optimal forecasts of GDP components are an essential part of an optimal overall economic growth forecast. Research in this field mostly focuses on predictions of private consumption (see, for instance, Vosen and Schmidt, 2011). Regarding trade developments, Ito (1990) finds behavioral biases in expectations of importers and exporters with respect to exchange rate changes, warranting a closer inspection of macroeconomic trade forecasts. Despite the important role of trade developments in macroeconomic forecasting, the evaluation of trade forecasts has received little attention in the academic literature.

Research on trade forecasts has mostly focused on their formation, where one strand of literature aims at modeling economic environments to simulate trade dynamics by means of large structural models (Hervé et al., 2011; Riad et al., 2012) and another strand of literature analyzes time series models in order to optimize their forecasting performance (Frale et al., 2010; Jakaitiene and Dées, 2012; Keck et al., 2009; Yu et al., 2008). Despite the increased public attention to German trade policy in recent years, studies with a focus on German trade are scarce. For German and Swiss export forecasts, Grossmann and Scheufele (2019) show that survey-based indicators improve forecast accuracy. Elstner et al. (2013) use manufacturing orders received for

[1] For a comparison of direct and disaggregated forecasting approaches, see, for instance, Angelini, Bańbura, and Rünstler (2010) and Heinisch and Scheufele (2018).

Germany, the real effective exchange rate, and an export expectations survey to forecast German exports. Jannsen and Richter (2012) predict German capital goods exports by establishing a capacity utilization indicator for German trading partners. Regarding German imports, Grimme, Lehmann, and Noeller (2018) introduce a leading indicator for total import growth based on the import demand of a country's main trading partners. Hetemäki and Mikkola (2005) compare the forecast accuracy of German paper import forecasts computed by means of several univariate time series models, single equation econometric models, multivariate systems models, and combinations thereof. The evaluation of trade forecasts has not received much attention in the literature, even though inefficient trade forecast can potentially be costly, if a protectionist trade policy is pursued based on such forecasts.

Several research institutes in Germany frequently publish economic forecasts. The evaluation of these forecasts is widespread and has received much attention in the scientific community. Kirchgässner (1993) and Sinclair, Stekler, and Müller-Dröge (2016) combine early approaches by Kirchgässner (1984) and Neumann and Buscher (1980) to rank several German research institutes based on the properties of their forecasts in a multivariate setting. Other studies focus on the effects of costly forecast revisions (Kirchgässner and Müller, 2006), or analyze the change of forecast accuracy over time (Heilemann and Stekler, 2013). Döpke and Fritsche (2006) use a panel-based approach to analyze growth and inflation forecasts (for a time series approach, see Kirchgässner and Savioz, 2001). Behrens, Pierdzioch, and Risse (2018a,c) use nonparametric tree-based models in univariate settings to analyze German GDP growth and inflation forecasts. Yet again, the evaluation of German trade forecasts has not received much attention, despite Germany's role as one of the largest exporters in the world and an increasing interest in Germany's trade policies in recent years. This study is a first attempt at closing this gap.

I focus on the evaluation of joint efficiency of trade forecasts for Germany in a multivariate setting, as both export and import volumes determine a country's net exports. Since net exports are often in the focus of political debates on the introduction of protectionist trade policies, it is crucial to consider both trade aggregates in a multivariate approach. Furthermore, macroeconomic aggregates, such as the exchange rate or the oil price, tend to influence export volumes as well as import volumes at the same time (Engelke et al., 2019). New techniques for studying the quality of a vector of forecasts in a multivariate setting have mushroomed in recent years. Sinclair, Stekler, and Carnow (2012) for instance, have adapted the Holden and Peel (1990) approach to weak and strong forecast efficiency to a multivariate setting. They use a VAR model on growth, inflation, and unemployment forecasts to analyze their bias and efficiency, and the Mahalanobis distance to assess forecast accuracy. In a multivariate setting, the Mahalanobis distance is a superior performance metric compared to the Euclidean distance as it captures correlations between the response variables (for further details on the Mahalanobis distance, see McLachlan, 1999). The Mahalanobis distance is used as performance metric for macroeconomic forecasts in several studies (see, e.g., Banternghansa and McCracken, 2009; Sinclair and Stekler, 2013; Sinclair, Stekler, and Carnow, 2015). Behrens, Pierdzioch, and Risse (2018b) incorporate the Mahalanobis distance in a nonparametric multivariate random forest model to study the joint efficiency of GDP growth and inflation forecasts. Multivariate tree-based models have first been introduced by Segal (1992) (see also Breiman, 2001; Segal and Xiao, 2011). I adapt the Behrens et al. (2018b) model to a trade setting, where I analyze the joint efficiency of trade forecasts for Germany. I use a novel data set, consisting of annual export and import forecasts for the years 1970 to 2017 with a forecast horizon of half-a-year and one year. Forecasts are available for four leading German economic research institutes, a collaboration of German forecasters, and one international forecaster. By means of these data, I compute multivariate random forests to model links between forecast errors and a forecaster's information set, which, in this study, consists of several trade and macroeconomic predictor variables. The reason I use a nonparametric tree-based approach is that linear forecasting models or evaluation techniques run into problems with the data at hand, as they exhibit relatively few and irregularly spaced observations, as well as possible nonlinearities among the predictor variables or between predictor and response variables.[2] A nonparametric approach overcomes resulting problems, such as a lack of degrees of freedom or model misspecification issues.

[2]For details on the analyzed data, see Section 3.

The remainder of this paper is structured as follows: I establish the concept of multivariate random forests in chapter 2. In chapter 3 the data are presented and chapter 4 outlines the empirical analysis. In chapter 5 I conclude.

2 Multivariate Random Forests

In order to motivate my nonparametric tree-based approach to test for joint forecast efficiency, I first ellaborate on the problems which may arise when using the widely recognized linear Mincer and Zarnowitz (1969) approach, before introducing multivariate random forests. Using a linear ordinary least squares (OLS) regression framework, Mincer and Zarnowitz define a forecast as efficient and unbiased if the realization of a given variable equals its forecast plus a residuum. However, several problems arise: First, the researcher has to define a linear function to model the relationship between the predictor and the forecast, a-priori, where the choice of predictor can be economically motivated but is still arbitrary to some degree. Second, with only one predictor, the implemented information set is quite narrow. One could argue to add further predictors, yet, this results in a loss of degrees of freedom and is especially problematic if a researcher only has a limited number of observations available, as is the case with the study at hand. Another option would be to reestimate the linear model with different combinations of predictors, which could quickly become a trial-and-error search. Furthermore, the relationship between the predictors and the forecasted variable might not always be linear. While this can, to some degree, be embedded in a linear regression framework, for instance, by means of squared predictors or indicator functions, again, the nonlinear relationship needs to be predefined by the researcher and can take on many more forms than quadratic or piecewise constant functions. Analogous assumptions need to be made if the researcher suspects an interdependency between two predictors. Once more, the two predictors could be linked multiplicatively in the OLS framework. However, this does not allow for a flexible interaction between the chosen predictors.

A way to overcome these drawbacks is to use a multivariate random forest model.[3] Roughly

[3]For a comprehensive introduction to tree-based models, see Hastie et al. (2009) and James et al. (2015). For

speaking, tree-based models let the data decide on an appropriate form to model interdependencies, rather than, a-priori, fitting a possibly misspecified model, which is not supported by the data. Tree-based models are nonparametric approaches that very broadly operate in a two step process.

1. The predictor space, $\mathbf{X}_t$, consisting of all possible combinations of the predictor variables, $\{x_{1t}, x_{2t}, ..., x_{it}\}$, is split into non-overlapping regions, R, using a subset of the analyzed dataset, the training data.

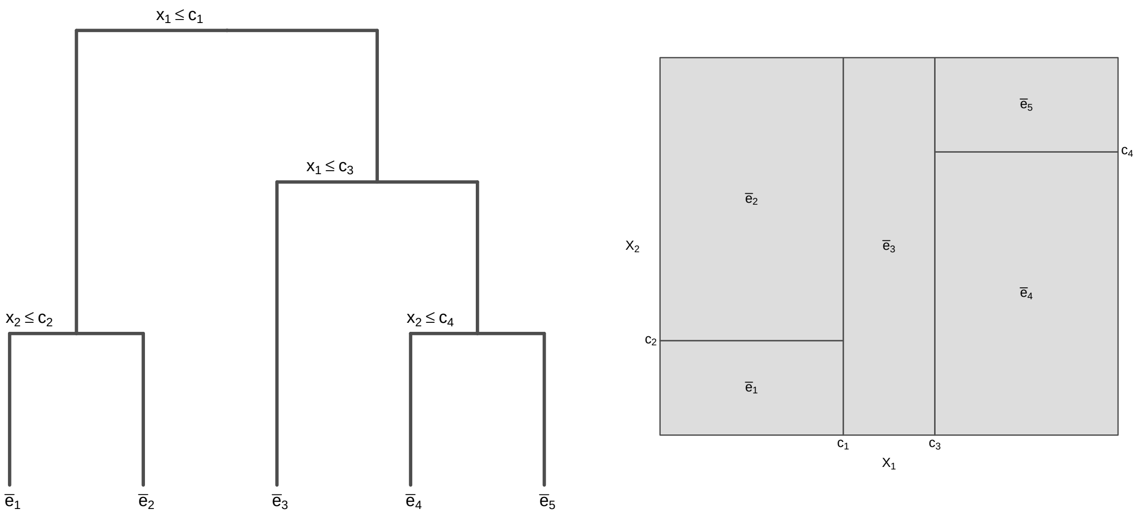
2. By means of the remaining test data, the so called out of bag data, a prediction is made for every observation based on the region to which it is assigned. The prediction is the mean of all response variables, i.e. the vectors of forecast errors, $\mathbf{e}_{t+1}$, in that region.

I now turn to the question how the predictor space is partitioned by the tree. To simplify the explanation of the basis of this recursive and binary process of region building, I first examine a univariate tree and use a notation similar to the one used by Hastie et al. (2009), where I drop the time index, t. In a first step, the whole data set is located at the top of the tree and all observations are part of one single region. Here, at the first node or root of the tree, similar to a standard least-squares regression, the residual sum of squares (RSS) is minimized by choosing a partitioning predictor, j, and a cut point, c, to split the dataset. This split creates two subsequent regions and observations are sent down the left branch to region one, R_1, if $j \leq c$ and down the right branch to region two, R_2, if $j > c$. Formally, the regions are built, such that $R_1(j,c) = \{x_j | x_j \leq c\}$ and $R_2(j,c) = \{x_j | x_j > c\}$ solve $\min_{j,c}\{RSS_1 + RSS_2\}$, where $RSS_m = \sum_{x_j \in R_m(j,c)}(e_i - \bar{e}_m)^2$, with $\bar{e}_m = \text{mean}\{e_i | x_j \in R_m(j,c)\}$, $m = \{1,2\}$, and e_i being a forecast error sent to region m. This process of partitioning the predictor space is continued until an a-priori defined stopping criterion is met. Stopping criteria are for instance a minimum number of response variables, i.e. forecast errors, at the terminal nodes or a maximum number of overall nodes. Figure 1 depicts this process

further details on random forests, see Breiman (2001) and on multivariate random forests, see Segal and Xiao (2011). For a numerical example of a multivariate random forest, see Behrens et al. (2018b).

of region building. In the left panel an exemplary tree is depicted. At each node, the dataset is divided into two further regions. At the root, data with partitioning predictors $x_1 > c_1$ are sent to the right child node, where another split occurs. Here, data with $x_1 \leq c_3$ are sent to the left, reaching a terminal node, or leaf, where the tree predicts a mean forecast error of $\bar{e}_3$. Data with $x_1 > c_3$ are sent to the right node, with the partitioning predictor x_2. If $x_2 > c_4$, $\bar{e}_5$ is the mean forecast error for these data, otherwise $\bar{e}_4$. Analogous, for data with $x_1 \leq c_1$, the prediction of the forecast error is $\bar{e}_1$ if $x_2 \leq c_2$ and $\bar{e}_2$ if $x_2 > c_2$. The panel on the right shows the predictor set which has been partitioned by this exemplary tree. It nicely depicts how a single tree fully describes the whole feature space. For all combinations of values of x_i, with $i = \{1,2\}$, the tree predicts the respective mean forecast error of each region.

Figure 1: Binary and Recursive Region Building Process of a Univariate Regression Tree



Note: Left panel: Exemplary univariate regression tree with partitioning predictors, x_t, split points, c_k, and mean response of forecast errors, $\bar{e}_i$. With $t = \{1,2\}, k = \{1:4\}, i = \{1:5\}$. Right panel: Partitioned feature space with the respective mean responses of forecast errors for each region. Example based on James et al. (2015, Figure 8.3).

I now adapt this process to a multivariate setting, in which the response variable is a vector, $\mathbf{e}_{t+1}$, of forecast errors, for instance computed from export and import growth forecasts. Hence, an appropriate RSS measure, which measures the distance of the vector of forecast errors, $\mathbf{e}_i$, sent to node m from the vector of the means of all response variables in region m, $\bar{\mathbf{e}}_m$.[4] To this end, I use the Mahalanobis distance as multivariate residual sum of squares (MRSS) measure as it appropriately weights the distance between $\mathbf{e}_i$ and $\bar{\mathbf{e}}_m$ to account for differences in the variances of the studied forecast errors and possible correlations (McLachlan, 1999). It is defined as $MRSS_m = \sum_{x_j \in R_m(j,c)} (\mathbf{e}_i - \bar{\mathbf{e}}_m)\mathbf{V}^{-1}(\mathbf{e}_i - \bar{\mathbf{e}}_m)'$, where $\mathbf{V}^{-1}$ is the variance-covariance matrix of the forecast errors.[5] When $\mathbf{V}^{-1}$ is the identity matrix, the Mahalanobis distance equals the Euclidean distance (see, for example, Sinclair, Stekler, and Müller-Dröge, 2016). Another advantage of the Mahalanobis distance is, that, besides exchanging MRSS for RSS, the computation of a multivariate regression tree is the same as in the univariate case. In fact, there are two differences between a multivariate and a univariate regression tree. The first one being the different measures of the residual sum of squares. The second difference is that a multivariate tree predicts a vector of response variables, i.e. $\mathbf{e}_{t+1}$, at the terminal nodes and a univariate tree predicts single responses, i.e. e_{t+1}.

The specific feature of a random tree is that for every split only a random subset of the elements of $\mathbf{X}_t$ is considered as partitioning predictors at each node of a tree. This characteristic decorrelates the predictions from individual trees and speeds up computations. However, a common problem with tree based models is that their potentially complex and hierarchical structure causes predictions of a single univariate or multivariate tree to be highly variable. In order to overcome this issue, I compute random forests, which use a large number of trees, grown independently from each other, to model the response variable (Breiman, 2001). I use bootstrap resampling and estimate a random tree on every bootstrapped dataset.[6] A further advantage of using a bootstrap is that it automatically generates out of bag test data, by means of which the performance of a random tree can be measured using the Mahalanobis distance as performance criterion.[7] In Section 4.1, I will elaborate on the Mahalanobis distance as key performance measure in my

[4]I again drop the time index, t, for simplification.

[5]For the case of a region-specific covariance matrix, see Segal (1992).

[6]For an introduction to bootstrap resampling, see for instance James et al. (2015).

[7]For an introduction to out of bag error estimation and an application to regression trees, see Hastie et al. (2009).

analysis.

3 The Data

I use a novel data set on annual export and import growth forecasts for Germany published by six economic research institutes. The list of forecasters comprises four German economic research institutes (alphabetical order): Deutsches Institut für Wirtschaftsforschung Berlin (DIW), Hamburgisches Weltwirtschaftsarchiv/-institut (HWWI), ifo Institut für Wirtschaftsforschung Munich (ifo), Institut für Weltwirtschaft Kiel (IfW); one collaboration of German economic research institutes: Gemeinschaftsdiagnose (GD); and the Organisation for Economic Co-operation and Development (OECD) as an international forecaster to check for a possible advantage due to geographical proximity of German forecast institutes (see for instance, Bae et al., 2008; Berger et al., 2009; Malloy, 2005). Forecasts are available for the years from 1970 to 2017, where the number of forecasts per year differs across forecasters and decades. I focus on annual forecasts with a forecast horizon of one year (1Y) and half-a-year (.5Y), due to data availability. The former are published at the turn of the year, whereas the latter are published mid-year. An exception are the forecasts formed by GD, which are published roughly two months before the other forecasts, namely in April and October. GD's forecasts, therefore, exhibit a longer forecast horizon than the other forecasts. In addition to the forecast data, I use initial release data from the German statistical office to measure realized values of export and import growth.[8] The German statistical office publishes such realized values for national accounts data of the previous year within the first months of a given year. In using these data, the effects of data revisions are minimized.

I subtract these realized values for export and import growth from the forecast for the respective year to compute the forecast error (see also, Behrens et al., 2018a), $e_{t(h),i,z} = \hat{y}_{t(h),i,z} - y_{t,z}$. Here, $\hat{y}_{t(h),i,z}$ denotes the $h = \{0.5, 1\}$-year-ahead annual forecast formed by forecaster i for $z = \{$export,

[8]Data taken from "Wirtschaft und Statistik" publications: `https://www.destatis.de/EN/Publications/WirtschaftStatistik/WirtschaftStatistik.html`.

import} for year $t = \{1970 : 2017\}$. After the German reunification the forecasters switch from forecasts for West-Germany to forecasts for reunified Germany at different dates between 1992 and 1993. I account for this by adjusting the corresponding time series for realized export and import growth for each forecaster.

Table 1: Descriptive Statistics of Forecast Errors

Forecaster	Horizon	*N*	MEAN	RMSE	RMAE	*N*	MEAN	RMSE	RMAE
		Exports				Imports			
DIW	***.5Y***	36	0.20	2.33	1.37	36	0.71	2.89	1.43
	1Y	48	-0.78	5.06	1.98	48	-0.28	3.95	1.76
HWWI	***.5Y***	31	-0.11	2.41	1.32	31	0.24	2.76	1.44
	1Y	43	-0.14	4.99	1.96	42	0.35	3.98	1.77
ifo	***.5Y***	41	-0.06	2.37	1.37	41	0.02	2.27	1.29
	1Y	44	-0.50	4.70	1.91	44	-0.09	3.48	1.66
IfW	***.5Y***	39	0.26	2.74	1.42	39	0.57	2.43	1.40
	1Y	47	-0.84	4.32	1.83	47	0.08	3.44	1.66
GD	***.5Y***	46	-0.58	3.77	1.72	47	-0.10	3.15	1.54
	1Y	48	-0.16	5.29	2.00	48	0.13	4.20	1.80
OECD	***.5Y***	46	0.38	3.19	1.58	46	0.29	3.51	1.60
	1Y	46	0.09	5.47	2.00	46	0.29	4.16	1.74

Notes: *N*: Number of observations. MEAN: Arithmetic mean. RMSE: Root-mean-squared error. RMAE: Root-mean-absolute error. .5Y: Half-a-year. 1Y: One year.

Table 1 shows the descriptive statistics for the forecast errors of all forecasters in the sample. Most observations ($N = 48$) are available for one-year-ahead forecasts of DIW and GD, whereas HWWI contributes the fewest observations ($N = 31$) for their half-a-year-ahead forecasts. The mean forecast errors are close to zero or at least smaller than one. Regarding the root mean squared error (RMSE) and root mean absolute error (RMAE) statistics, the results are, as one could have anticipated, generally larger for the one-year-ahead forecasts. Compared to mean forecast errors, RMSEs, and RMAEs usually obtained in the more common analyses of German inflation and growth forecasts, the depicted statistics are relatively high (see, for instance, Behrens, Pierdzioch, and Risse, 2018a; Döpke, Fritsche, and Pierdzioch, 2017). This is most probably due to the fact that trade aggregates tend to be rather volatile and therefore harder to predict. Döhrn and Schmidt (2011), for instance, analyze German macroeconomic forecasts and find that exports, imports, and investments are the most volatile components of GDP.

Table 2: Predictors

Predictors	Group	ln	Lag	Description	Source
Production Germany	M - Real	Y	1	Year-on-year rate of change of the monthly German total manufacturing output.	OECD
Production United States	M - Real	Y	1	Year-on-year rate of change of the monthly U.S. total manufacturing output.	OECD
Production France	M - Real	Y	1	Year-on-year rate of change of the monthly French total manufacturing output.	OECD
Production United Kingdom	M - Real	Y	1	Year-on-year rate of change of the monthly U.K. total manufacturing output.	OECD
Production Italy	M - Real	Y	1	Year-on-year rate of change of the monthly Italian total manufacturing output.	OECD
Production Netherlands	M - Real	Y	1	Year-on-year rate of change of the monthly Dutch total manufacturing output.	OECD
Order inflow	M - Real	Y	1	Year-on-year rate of change of the industrial orders received for Germany; calendar and seasonally adjusted.	BUBA
Unemployment	M - Real	N	1	Monthly unemployment rate (in percent of civilian labor) for Germany; calendar and seasonally adjusted.	BUBA
Oil price	M - Prices	Y	0	Year-on-year rate of change of the monthly crude oil price (WTI); dollars per barrel.	FRED
Climate	M - Survey	N	0	Monthly ifo business tendency survey for manufacturing for Germany; half-a-year-ahead tendency, seasonally adjusted.	FRED
Climate (expectations)	M - Survey	N	0	Monthly ifo business tendency survey for manufacturing for Germany; situation in six months; seasonally adjusted.	FRED
OECD leading (normalized)	M - Composite	N	2	Monthly normalized OECD composite leading indicator for Germany.	OECD
Real effective exchange rate	T - Prices	Y	1	Year-on-year rate of change of the monthly narrow effective exchange rate for Germany; CPI-based.	BIS
Exports	T - Real	Y	12	12-months-lag of the year-on-year rate of change of German value goods exports; seasonally adjusted.	FRED
Imports	T - Real	Y	12	12-months-lag of the year-on-year rate of change of German value goods imports; seasonally adjusted.	FRED
Export Prices	T - Prices	Y	1	Year-on-year rate of change of the monthly index of German export prices; standard international trade classification.	DESTATIS
Import Prices	T - Prices	Y	1	Year-on-year rate of change of the monthly index of German import prices; standard international trade classification.	DESTATIS
Consumer Prices	M - Prices	Y	0	Year-on-year rate of change of the monthly German consumer price index; calendar and seasonally adjusted.	BUBA
Producer Prices	M - Prices	Y	0	Year-on-year rate of change of the monthly German domestic producer price index for manufacturing.	FRED

Notes: BIS - Bank for International Settlements, `https://www.bis.org/statistics/index.htm`; BUBA - German Central Bank, `http://www.bundesbank.de/Navigation/EN/Statistics/statistics.html`; DESTATIS - Federal Statistical Office of Germany, `https://www.destatis.de/EN/FactsFigures/FactsFigures.html`; FRED - Federal Reserve Bank of St. Louis, `https://fred.stlouisfed.org/`; OECD - Organisation for Economic Co-operation and Development, `http://stats.oecd.org/`. ln: natural logarithmic transformation, Y - yes, N - no, M - Macroeconomic variable, T - Trade variable. – Lag: Publication lags in months added where necessary (Drechsel and Scheufele, 2012, table 3).

Table 2 summarizes the predictors I use to model a forecaster's information set at the time of forecast formation. A recent study of euro-area trade by D'Agostino et al. (2017) has shown that accounting for the interaction of macroeconomic and trade variables is essential to accurately predict future trade developments. I follow their approach and include, in addition to several macroeconomic variables, trade variables, namely export and import prices for Germany, German export and import volumes lagged by one year, as well as the German real effective exchange rate, which serves as a measure of international price competitiveness (see also Grimme et al., 2018; Lehmann, 2015). Consistent with recent research on German growth and inflation forecasts (Behrens et al., 2018c; Döpke et al., 2017) my list of macroeconomic predictors to assess the state of the German economy comprises German industrial orders, German consumer and producer price indices, the German unemployment rate, industrial production for Germany, and the oil price. I furthermore add industrial production for Germany's main trading partners since 1970, i.e. the United States, France, the United Kingdom, Italy, and the Netherlands, as a leading indicator of the respective country's economic development (see, e.g., Guichard and Rusticelli, 2011, who show that industrial production indices can improve trade forecasts).[9] Due to possible delayed effects of a change in the real effective exchange rate, I add the last four lags of the German real effective exchange rate as predictors (for evidence of a J-curve effect for German trade partners, see Hacker and Hatemi-J, 2004). Following Frale et al. (2010) and Lehmann (2015), I also use macroeconomic survey data as predictors, namely German business tendency surveys for manufacturing (current and future tendency). As a composite indicator, I include the normalized OECD leading indicator for Germany. The predictors are available on a monthly basis. I account for a publication lag of the forecasts by using predictors lagged by one period. In other words, if a forecast was published in, for instance, December, I assume that it is based on predictors available in November (regarding the publication lags of the predictors, I follow Drechsel and Scheufele, 2012). As aforementioned, I model the information available to a forecaster when he or she formed a particular forecast. I therefore follow Behrens et al. (2018c) and use a backward looking moving-average of order 12 to minimize data revision effects on the

[9] Due to data availability and the reason that China has only fairly recently become one of Germany's main trading partners, I do not include China's industrial production in my list of predictors.

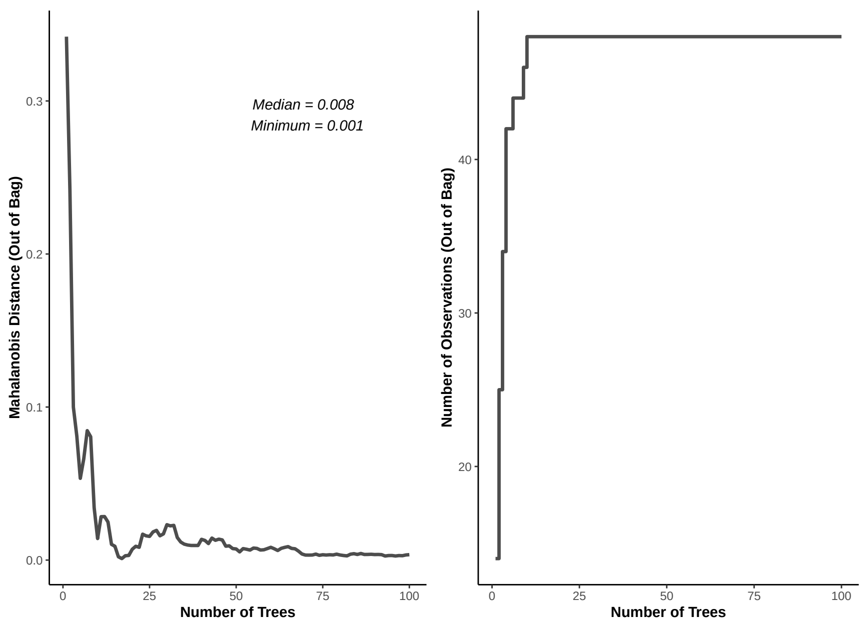
CPI, PPI, the real effective exchange rate, industrial production, orders, and trade variables, as for these predictors only revised time series are available (see also, Pesaran and Timmermann, 1995).

4 Empirical Analysis

4.1 The Model

For computations I use the R Core Team (2018) programming environment for statistical computing and the add-on package "MultivariateRandomForest" (Rahman, Otridge, and Pal, 2017). Regarding the parameters of the model, I follow Behrens et al. (2018b) and adapt the common approach in the relevant literature (see, e.g., Hastie et al., 2009) in setting the number of predictors used for building a random forest to the square root of the number of total predictors, and I set the number of terminal nodes to five observations.

Figure 2: Out of Bag Data in Random Forest



Note: The figure illustrates out of bag predictions of an exemplary random forest for the half-a-year-ahead forecasts of GD, computed by means of bootstrapped data.

The process of building a random forest consists, in a first step, of growing a series of $\{1{:}100\}$ single random trees. Each tree is grown from approximately two thirds of a bootstrapped data set. Such a tree is subsequently used to compute predictions about forecast errors in the remaining third of the bootstrapped data, i.e. the out of bag data. The plot in the left panel of figure 2 depicts for the one-year-ahead forecasts of GD a series of $\{1{:}100\}$ random trees. It visualizes the reason for growing a random forest instead of working with a single tree. In adding more trees to the random forest, the high variability of the Mahalanobis distance, when using only a minor number of trees, is reduced until it fluctuates around a low level. The plot on the right illustrates that, in the process of growing a random forest, every data point is used at least once as an out of bag prediction. Here, the $N = 48$ one-year-ahead forecasts of GD are depicted and when a random forest consists of about 10 random trees, every observation has been used as an out of bag prediction. Finally, the generated out of bag data are used to analyze the performance of a random forest, which I measure by means of the out of bag Mahalanobis distance, measuring the distance between the predicted forecast error and the actual forecast error. As opposed to the Euclidean distance, the Mahalanobis distance captures possible correlations between the response variables in a multivariate analysis (McLachlan, 1999). Finally, I use the Mahalanobis distance to infer the joint efficiency of the analyzed trade forecasts. Obtaining the Mahalanobis distances in my analysis follows a three-way process:

1. In building a random forest, the out of bag predictions of the single random trees are averaged over all trees as the random forest grows.
2. The Mahalanobis distance is used to measure the distance between the predicted forecast errors computed in step one and the actual forecast errors. In doing so, a series of $\{1{:}100\}$ Mahalanobis distances is recorded as the random forest grows.
3. Finally, the median and minimum of the series of Mahalanobis distances obtained in step two is used as accuracy criterion.

The computation of the Mahalanobis distance is executed by means of a bias adjusted variance-covariance matrix of the predicted and actual forecast error, $\mathbf{V} = \frac{(\mathbf{V}_{\hat{e}_{t+1}} + \mathbf{V}_{e_{t+1}})(N-1)}{(N-2)}$, where N

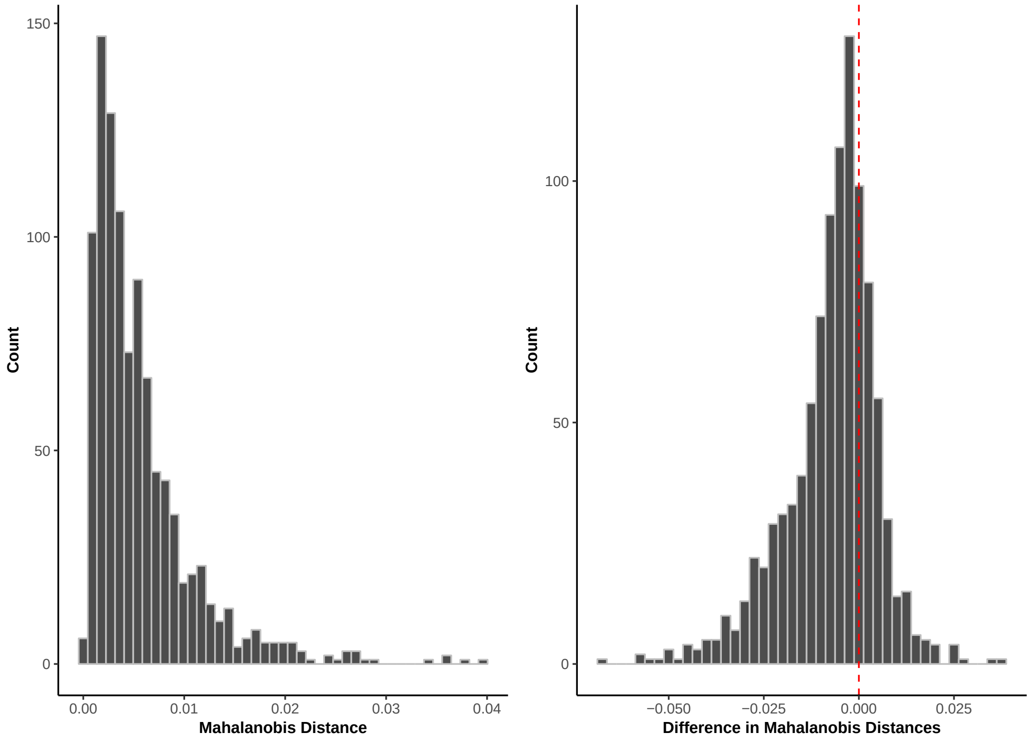
is the number of observations, $\mathbf{V}$ is the pooled variance-covariance matrix of the bias-corrected variance-covariance matrices of $\hat{e}_{t+1}$ and e_{t+1}, with $\hat{e}_{t+1}$ denoting the averaged out of bag forecasts in a random forest obtained above in step one (see also, Behrens et al., 2018b).

To account for the fact that the computed median and minimum Mahalanobis distances are random numbers, the aforementioned steps are repeated $D = 1{,}000$ times. In order to asses the informative value of the obtained Mahalanobis distances, I run a permutation test by computing a second series of Mahalanobis distances, which should not have any predictive value for the forecast errors, and comparing these Mahalanobis distances with the ones computed by means of the original data. The former Mahalanobis distances are obtained from estimating a multivariate random forest 1,000 times on a permuted data set, where the matrix of forecast errors is permuted by sampling without replacement from the original data. Under the null hypothesis of joint forecast efficiency, the set of predictors should neither have predictive power for the permuted forecast errors nor for the corresponding, original forecast errors. In other words, under the null hypothesis, the Mahalanobis distance computed by means of estimating a multivariate random forest on the original data should not be smaller than the Mahalanobis distance computed by means of estimating a multivariate random forest on the permuted data.

In a final step, I test for the statistical significance of the difference between the Mahalanobis distances, ΔM, by regressing it on a constant: $\Delta M_i = \beta + \mu_i$, with $i = \{1 : D\}$ and μ_i being an error term. The difference between the Mahalanobis distances is computed as $\Delta M_i = M_i^O - M_i^P$, where M_i^O is the $D \times 1$ vector of Mahalanobis distances estimated from the original data and M_i^P is the $D \times 1$ vector of Mahalanobis distances estimated from the permuted data. I therefore reject the null hypothesis of joint forecast efficiency, if the estimated coefficient, $\hat{\beta}$, is significantly negative. The estimates of such a constant are computed by means of an OLS regression model, as well as a robust linear regression model (RLM) using Huber's loss function. The latter model is used as a robustness check in accounting for the tail behavior of the sampling distribution of ΔM.[10]

[10] I follow Behrens et al. (2018b) in this approach and use the R add-on package "MASS" for computations (Venables and Ripley, 2002).

Figure 3: Derivation of the Sampling Distribution of ΔM for IfW (1Y-Forecasts)



Note: $\Delta M = M^O - M^P$, i.e., ΔM is the difference between the median Mahalanobis distance computed by means of the original data and the Mahalanobis distance computed by means of the permuted data.

The derivation of the sampling distribution of ΔM for IfW's one-year-ahead trade forecasts is exemplarily depicted in Figure 3. From a series of $\{1{:}100\}$ Mahalanobis distances, similar to the ones shown in the left panel of Figure 2, I select the median and minimum distance, which are both random numbers. Next, the process is repeated 1,000 times and, as an example, the left panel of Figure 3 illustrates the obtained sampling distribution of the median Mahalanobis distances for IfW. Its shape is common for Mahalanobis distances, as they are distributed as non-central F-distributions (McLachlan, 1999). The right panel shows the sampling distribution of the differences between the Mahalanobis distances computed from the original and the permuted data. In order to account for negative skewness and excess kurtosis of the distribution, I use ordinary OLS models as well as robust RLM models, when analyzing the mean of ΔM.

Furthermore, negative skewness of the distribution is to be expected in the case that forecasts are not jointly efficient. This is because if forecasts are jointly inefficient, at least one predictor has predictive power, resulting, over multiple simulations, in a significantly shorter Mahalanobis distance estimated on the original data relative to the one estimated on the permuted data. The results presented in section 4.2 confirm this train of thought, as I find significant evidence against the joint efficiency of IfW's one-year-ahead trade forecasts.

4.2 Results

Table 3 depicts the results of the model's basic specification. Bold numbers indicate significance for a negative test statistic at least at the 5%-level. I only reject joint forecast efficiency when the estimated coefficient is negative, the intuition being that a small Mahalanobis distance indicates predictive power of at least one predictor. Hence, the Mahalanobis distance retrieved from the original data should, on average, be smaller than the Mahalanobis distance retrieved from the permuted data, producing a negative estimate of the coefficient and a rejection of joint forecast efficiency.

I strongly reject joint efficiency of forecasts for DIW and IfW. In both cases, the estimated coefficients for the half-a-year-ahead as well as the one-year-ahead forecasts are negative and significant for the OLS and the RLM models, using both, the median and minimum Mahalanobis distances. For ifo, on the other hand, I cannot reject joint forecast efficiency, as the coefficients (.5Y- and 1Y-forecasts) and the test statistics (1Y-forecasts) are positive for all specifications. Regarding HWWI, I cannot reject joint efficiency for the half-a-year-ahead forecasts. The estimated coefficients are equal to zero and insignificant. The one-year-ahead forecasts, however, appear to be jointly inefficient. In turn, for GD and OECD, I cannot reject joint efficiency of the 1Y-forecasts, whereas joint efficiency of the .5Y-forecasts is rejected. An exception is the coefficient of the RLM model for GD, computed using the minimum Mahalanobis distance, which is insignificant.

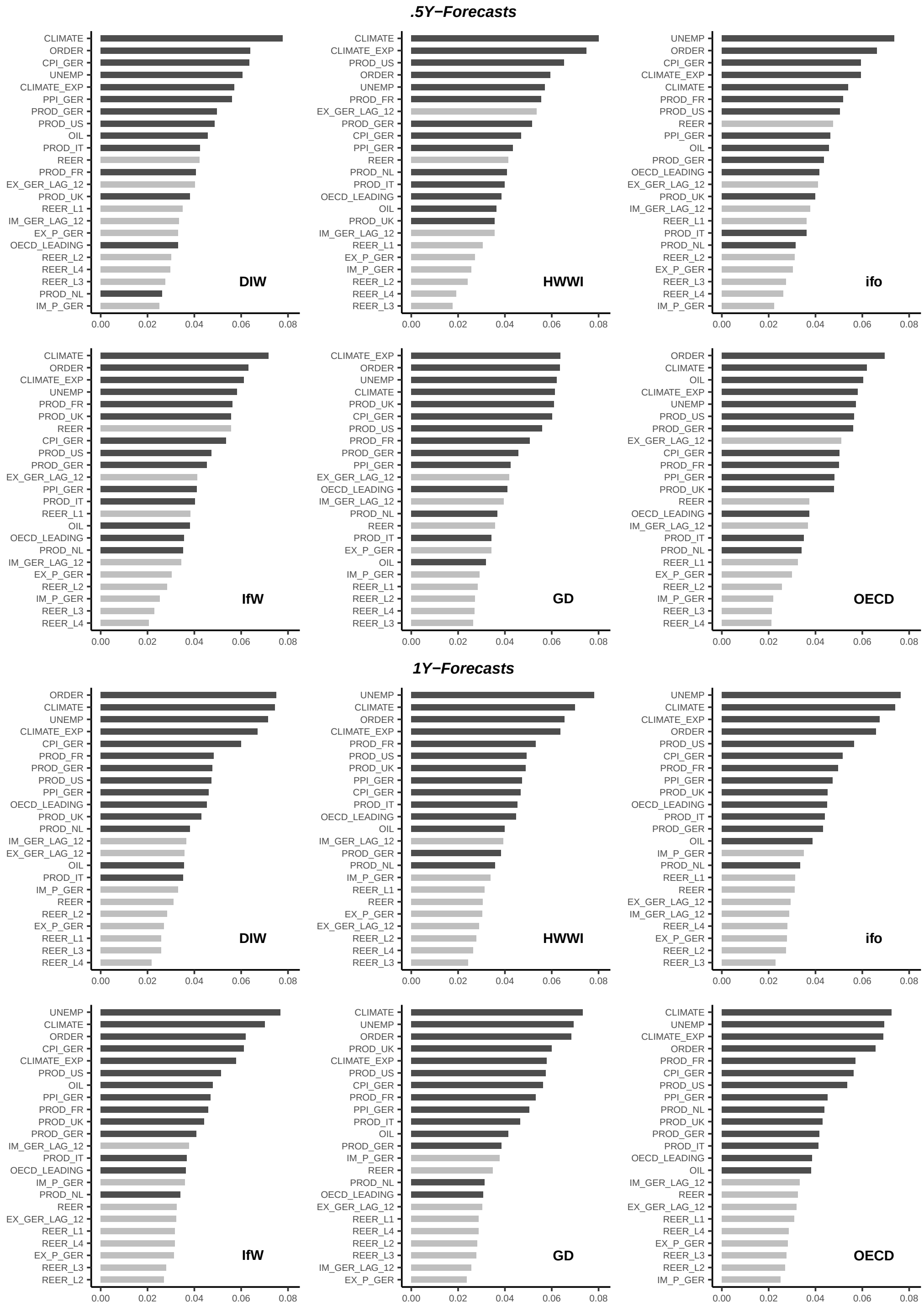
Table 3: Basic Specification

Forecaster	Horizon	β_{OLS}	t_{OLS}	β_{RLM}	t_{RLM}	β_{OLS}	t_{OLS}	β_{RLM}	t_{RLM}
		Median				**Minimum**			
DIW	**.5*Y***	**-0.005**	**-11.291**	**-0.004**	**-10.325**	**-0.002**	**-9.492**	**-0.001**	**-8.799**
	1*Y*	**-0.002**	**-5.793**	**-0.002**	**-5.777**	**-0.001**	**-3.019**	**0.000**	**-2.145**
HWWI	**.5*Y***	0.000	-0.130	0.000	0.371	0.000	0.173	0.000	0.936
	1*Y*	**-0.005**	**-10.818**	**-0.004**	**-9.885**	**-0.001**	**-6.848**	**-0.001**	**-5.067**
ifo	**.5*Y***	0.004	7.514	0.004	9.118	0.002	5.399	0.001	7.814
	1*Y*	0.000	0.365	0.001	1.457	0.000	1.112	0.000	2.944
IfW	**.5*Y***	**-0.006**	**-13.503**	**-0.004**	**-12.578**	**-0.002**	**-9.768**	**-0.001**	**-9.666**
	1*Y*	**-0.007**	**-18.181**	**-0.006**	**-18.625**	**-0.002**	**-12.681**	**-0.001**	**-13.283**
GD	**.5*Y***	**-0.002**	**-3.818**	**-0.001**	**-2.514**	**-0.001**	**-3.185**	0.000	-0.978
	1*Y*	0.002	4.534	0.002	5.881	0.000	2.129	0.001	3.809
OECD	**.5*Y***	**-0.005**	**-15.177**	**-0.004**	**-14.743**	**-0.002**	**-11.091**	**-0.001**	**-10.346**
	1*Y*	0.004	7.986	0.004	10.060	0.002	7.730	0.002	9.285

Notes: Median: Results of permutation tests computed by means of the median of the Mahalanobis distance estimated for every random forest consisting of 100 random trees. Minimum: Results of permutation tests computed by means of the minimum of the Mahalanobis distance estimated for every random forest consisting of 100 random trees. Results are based on the computation of 1,000 random forests. OLS: ΔM is regressed onto a constant and results are obtained by an ordinary-least squares regression. RLM: ΔM is regressed onto a constant and robust results are obtained by a robust regression using Huber's loss function. $\Delta M = M^O - M^P$, i.e., ΔM is the difference between the median Mahalanobis distance computed by means of the original data and the Mahalanobis distance computed by means of the permuted data. β: Estimated coefficient. t: t-test. .5Y: Half-a-year. 1Y: One year. Boldface numbers indicate significance at least at the 5% level (for a negative test statistic).

In order to add further economic intuition to my analysis, I measure the importance of a certain variable in growing a random forest by counting the times a predictor is selected as splitting variable in a regression tree. I then sum up the variable importance measures of each tree and compute the share of each predictor as a splitting variable in a random forest. This percentage share is averaged over all bootstrapped random forests and presented in Figure 4. A high relative importance indicates explanatory power of a certain predictor for the forecast error. This implies that said predictor is not efficiently incorporated in the respective institute's trade forecast. In turn, a low relative importance indicates that the respective predictor has no explanatory power for the forecast error, which implies that either the predictor has been efficiently incorporated in an institutes trade forecast or that the respective variable generally has no predictive power in a trade forecast. The variable importance plots show that the typical trade predictors in the information set are more rarely selected as splitting variables than other macroeconomic vari-

Figure 4: Variable Importance Plots



Notes: Variable importance is measured by counting the times a predictor is selected as splitting variable in a regression tree. Variable importance measures of all trees are summed up and the share of each predictor as a splitting variable in a random forest is computed. This percentage share is averaged over all bootstrapped random forests. Light grey bars indicate trade variables, dark grey bars indicate other macroeconomic variables.

ables. Since the multivariate random forest randomly picks a subset of predictors from which to choose to minimize the node impurity measure, i.e. the Mahalanobis distance, every predictor has a certain share of being selected as splitting variable.

For all forecasters, most trade variables exhibit a relative importance of less than 4%, whereas the top predictors are chosen as splitting variables twice as often. In the trade variable category, the 4%-mark is exceeded only by the real effective exchange rate and lagged German export volumes for the half-a-year-ahead forecasts and by the lagged German import volumes for the one-year-ahead forecasts. Among the top predictors, i.e. the predictors which are not used efficiently by the forecasters and, therefore, contain information which can reduce forecast errors, are exclusively macroeconomic variables. Especially survey indicators, such as the ifo business climate survey, real macroeconomic indicators, such as the unemployment rate and industrial orders, as well as price indicators, such as the consumer price index, seem to contain useful information to reduce the forecast error for all forecasters and forecast horizons. Overall the variable importance plots suggest that the forecasters use typical trade indicators efficiently, whereas other macroeconomic predictors, receive insufficient attention. This observation lends support to the aforementioned study of D'Agostino et al. (2017) who find for the Euro-Zone that, besides trade variables, macroeconomic variables are important to accurately predict trade developments.

Furthermore, it is interesting to see whether OECD and ifo use the OECD leading indicator and the ifo business climate index more efficiently due to possible information asymmetries. However, the variable importance plots suggest that neither OECD nor ifo use the respective predictors more or less efficiently than the other forecasters. For all institutes and for both forecast horizons, the ifo business climate index is among the top predictors. The OECD leading indicator is a middle-ranking predictor for all institutes and forecast horizons.

In order to account for possible sticky-information processing or rational inattention in a forecaster's information set (see, e.g., Andrade and Le Bihan, 2013, on inattentive professional forecasters), I add the lagged realizations of the predictors to the set of predictors of the basic specification and report the results in Table 4.

Table 4: Extended Information Set

Forecaster	Horizon	β_{OLS}	t_{OLS}	β_{RLM}	t_{RLM}	β_{OLS}	t_{OLS}	β_{RLM}	t_{RLM}
		Median				**Minimum**			
DIW	**.5*Y***	**-0.003**	**-6.052**	**-0.002**	**-5.265**	**-0.001**	**-4.546**	**0.000**	**-2.395**
	1*Y*	**-0.002**	**-4.212**	**-0.001**	**-2.817**	**-0.001**	**-2.509**	0.000	-0.008
HWWI	**.5*Y***	**-0.002**	**-3.446**	**-0.001**	**-2.148**	**-0.001**	**-2.851**	0.000	-0.503
	1*Y*	**-0.006**	**-15.319**	**-0.005**	**-14.926**	**-0.002**	**-10.812**	**-0.001**	**-10.484**
ifo	**.5*Y***	0.002	3.616	0.002	5.373	0.001	5.092	0.001	7.257
	1*Y*	**-0.001**	**-2.343**	0.000	-0.952	0.000	-0.888	0.000	0.718
IfW	**.5*Y***	**-0.005**	**-10.704**	**-0.004**	**-10.337**	**-0.001**	**-7.266**	**-0.001**	**-6.900**
	1*Y*	**-0.008**	**-19.647**	**-0.007**	**-19.338**	**-0.003**	**-14.679**	**-0.002**	**-14.144**
GD	**.5*Y***	**-0.003**	**-6.608**	**-0.002**	**-6.366**	**-0.001**	**-5.128**	**-0.001**	**-4.671**
	1*Y*	0.003	6.945	0.003	8.081	0.001	4.804	0.001	7.143
OECD	**.5*Y***	**-0.005**	**-15.019**	**-0.004**	**-13.942**	**-0.002**	**-10.750**	**-0.001**	**-10.421**
	1*Y*	0.005	9.744	0.005	12.887	0.002	7.795	0.002	10.374

Notes: Median: Results of permutation tests computed by means of the median of the Mahalanobis distance estimated for every random forest consisting of 100 random trees. Minimum: Results of permutation tests computed by means of the minimum of the Mahalanobis distance estimated for every random forest consisting of 100 random trees. Results are based on the computation of 1,000 random forests. OLS: ΔM is regressed onto a constant and results are obtained by an ordinary-least squares regression. RLM: ΔM is regressed onto a constant and robust results are obtained by a robust regression using Huber's loss function. $\Delta M = M^O - M^P$, i.e., ΔM is the difference between the median Mahalanobis distance computed by means of the original data and the Mahalanobis distance computed by means of the permuted data. β: Estimated coefficient. t: t-test. .5Y: Half-a-year. 1Y: One year. Boldface numbers indicate significance at least at the 5% level (for a negative test statistic).

There is still evidence against the joint efficiency of half-a-year-ahead and one-year-ahead forecasts of DIW, however, the result for the latter using the minimum Mahalanobis distance for the RLM specification becomes insignificant. The results for ifo change for the 1Y-forecast coefficient, estimated by means of the median Mahalanobis distance and an OLS regression, which is now significantly negative. For all other forecasts and specifications of ifo, I cannot reject joint forecast efficiency. In addition to the one-year-ahead forecasts of HWWI, I reject, on the basis of the extended information set, the joint efficiency of their half-a-year-ahead forecasts as well. An exception is the RLM estimate for the minimum Mahalanobis specification, which is insignificant. The results of GD and OECD are also qualitatively close to the ones obtained in the basic specification of the model. There is evidence against the joint efficiency of their half-a-year-ahead forecasts, where the estimated coefficient for the RLM specification, using the minimum Mahalanobis distance of GD is now significantly negative. Finally, for IfW, the results also do not change qualitatively, I strongly reject the joint efficiency of forecasts for both forecast horizons and all regression model specifications. What to take away from these results is, that for most forecasters there is no major change in the results, when extending the assumed information set by adding lagged realizations of the predictor variables. Only for HWWI, there seems to be evidence that the lagged predictors have additional predictive power for the half-a-year-ahead forecast errors.

Table 5: Without Financial Crisis

Forecaster	**Horizon**	β_{OLS}	t_{OLS}	β_{RLM}	t_{RLM}	β_{OLS}	t_{OLS}	β_{RLM}	t_{RLM}
		Median				**Minimum**			
DIW	**.5*Y***	**-0.006**	**-12.297**	**-0.005**	**-12.236**	**-0.002**	**-8.859**	**-0.001**	**-8.481**
	1*Y*	**-0.003**	**-6.349**	**-0.002**	**-5.278**	**-0.001**	**-4.084**	**0.000**	**-2.829**
HWWI	**.5*Y***	0.006	8.620	0.006	9.692	0.002	6.678	0.002	7.264
	1*Y*	**-0.006**	**-14.402**	**-0.005**	**-13.215**	**-0.002**	**-10.093**	**-0.001**	**-8.479**
ifo	**.5*Y***	0.002	4.139	0.003	5.187	0.001	3.420	0.001	4.072
	1*Y*	**-0.001**	**-2.429**	0.000	-1.237	0.000	-0.457	0.000	1.511
IfW	**.5*Y***	**-0.006**	**-13.090**	**-0.005**	**-13.645**	**-0.002**	**-9.863**	**-0.001**	**-9.803**
	1*Y*	**-0.008**	**-17.898**	**-0.006**	**-17.937**	**-0.003**	**-13.085**	**-0.001**	**-12.284**
GD	**.5*Y***	**-0.001**	**-3.157**	**-0.001**	**-2.161**	**0.000**	**-1.735**	0.000	-0.402
	1*Y*	0.003	6.606	0.003	7.904	0.001	4.290	0.001	5.188
OECD	**.5*Y***	**-0.005**	**-13.919**	**-0.004**	**-13.938**	**-0.002**	**-10.876**	**-0.001**	**-10.220**
	1*Y*	0.004	8.980	0.005	11.602	0.002	6.528	0.002	9.754

Notes: Median: Results of permutation tests computed by means of the median of the Mahalanobis distance estimated for every random forest consisting of 100 random trees. Minimum: Results of permutation tests computed by means of the minimum of the Mahalanobis distance estimated for every random forest consisting of 100 random trees. Results are based on the computation of 1,000 random forests. OLS: ΔM is regressed onto a constant and results are obtained by an ordinary-least squares regression. RLM: ΔM is regressed onto a constant and robust results are obtained by a robust regression using Huber's loss function. $\Delta M = M^O - M^P$, i.e., ΔM is the difference between the median Mahalanobis distance computed by means of the original data and the Mahalanobis distance computed by means of the permuted data. β: Estimated coefficient. t: t-test. .5Y: Half-a-year. 1Y: One year. Boldface numbers indicate significance at least at the 5% level (for a negative test statistic).

Next, I consider the possibility that the financial crisis of 2007/2008 affects my results. In recent years, a large strand of literature has analyzed the effects of the 2007/2008 financial crisis on forecast formation and evaluation (see, e.g., Drechsel and Scheufele, 2012; Frenkel, Lis, and Rülke, 2011). I follow Behrens et al. (2018c), by excluding forecasts for the years 2007 and 2008. The intuition here is, that relatively large forecast errors produced in these years might bias the results of the whole sample.[11] However, this does not seem to be a problem with the sample at hand, as the results in Table 5 are very similar to the ones of the basic specification. In fact, only ifo's estimated coefficient of the 1Y-forecast using the median Mahalanobis distance and the OLS model is affected in becoming significantly negative. Regarding all other forecasters, the results do not change qualitatively. I still reject joint forecast efficiency for the half-a-year-ahead and one-year-ahead forecasts of DIW and IfW. The OLS and RLM estimates are significantly negative, showing that the (median and minimum) Mahalanobis distance computed by means of the multivariate random forests estimated on the original data is, on average, significantly smaller than the Mahalanobis distance computed by means of the multivariate random forests estimated on the permuted data. For ifo, I cannot reject joint forecast efficiency in all but the aforementioned case. With respect to HWWI, I continue to reject joint efficiency of their one-year-ahead forecasts. Finally, there remains to be evidence against the joint forecast efficiency of the .5Y-forecasts of GD and OECD, with one exception (.5Y-RLM-estimate using the minimum Mahalanobis distance for GD).

I conclude my analysis of joint forecast efficiency of the six forecasters, by compiling an efficiency ranking, reported in Table 7. In a similar way, Sinclair et al. (2016) and Behrens et al. (2018b) use the Mahalanobis distance to compute a ranking for forecast accuracy and efficiency, respectively. The rank of a forecaster is determined by summing up the t-statistics from the model specifications, reported in Tables 3-5. The intuition behind this is, that a small negative or positive t-statistic is evidence against a rejection of joint forecast efficiency, whereas a large negative t-statistic leads to a rejection of joint forecast efficiency. Hence, the sum of a given forecaster's t-statistics over all three scenarios is an indicator of the strength of evidence against joint forecast efficiency of said forecaster. Summing up the t-statistics implies that a positive t-statistic in one specification can compensate the effect of a negative t-statistic in another specification. Overall, the results are robust to changes in the regression model specification (OLS/RLM) and

[11] In the same vein, I check for a possible bias due to large forecast errors in the time of German reunification. The results for a sample, in which the forecasts for the years, in which the forecasters switch from forecasts for West- to reunified Germany, are excluded, are hardly affected. The results are not reported but available from the author upon request.

to changes in computations using the median or minimum Mahalanobis distance. However, for half-a-year-ahead and one-year-ahead forecasts, the results differ. Regarding the former forecast horizon, ifo and HWWI lead the ranking with a positive sum of t-statistics. OECD, least efficient in the .5Y-ranking, leads the ranking for the one-year-ahead forecasts followed by GD, both with positive sums of t-statistics. DIW and IfW are in the lower ranks for the 1Y- as well as the .5Y-forecasts. Based on these results, I cannot comment on a possible advantage due to geographical proximity of German forecasting institutes, as the OECD's rank drastically differs, depending on the examined forecast horizon.

Table 6: Forecaster Ranking

Rank	**Horizon**	t_{OLS}	t_{RLM}	t_{OLS}	t_{RLM}
		Median		**Minimum**	
I	*.5Y*	*ifo* (15.269)	*ifo* (19.678)	*ifo* (13.911)	*ifo* (19.143)
II		*HWWI* (5.044)	*HWWI* (7.915)	*HWWI* (4.000)	*HWWI* (7.697)
III		*GD* (-13.583)	*GD* (-11.041)	*GD* (-10.048)	*GD* (-6.051)
IV		*DIW* (-29.64)	*DIW* (-27.826)	*DIW* (-22.897)	*DIW* (-19.675)
V		*IfW* (-37.297)	*IfW* (-36.560)	*IfW* (-26.897)	*IfW* (-26.369)
VI		*OECD* (-44.115)	*OECD* (-42.623)	*OECD* (-32.717)	*OECD* (-30.987)
I	*1Y*	*OECD* (26.710)	*OECD* (34.549)	*OECD* (22.053)	*OECD* (29.413)
II		*GD* (18.085)	*GD* (21.866)	*GD* (11.223)	*GD* (16.140)
III		*ifo* (-4.407)	*ifo* (-0.732)	*ifo* (-0.233)	*ifo* (5.173)
IV		*DIW* (-16.354)	*DIW* (-13.872)	*DIW* (-9.612)	*DIW* (-4.982)
V		*HWWI* (-40.539)	*HWWI* (-38.026)	*HWWI* (-27.753)	*HWWI* (-24.030)
VI		*IfW* (-55.726)	*IfW* (-55.900)	*IfW* (-40.445)	*IfW* (-39.711)

Notes: In order to compile the ranking of forecasters, the t-statistics given in Tables 3–5 are added up and the resulting values are reported in parentheses. .5Y: Half-a-year. 1Y: One year.

5 Conclusion

I contribute to research on the evaluation of the properties of German macroeconomic forecasts. I build on research on the evaluation of forecasts in a multivariate setting (see, for instance, Sinclair et al., 2012; Sinclair and Stekler, 2013; Sinclair et al., 2015). As opposed to the majority of studies in this field, which analyze GDP growth and inflation forecasts, I study the joint efficiency of trade forecasts. To do so, I use multivariate random forests, an approach brought forward by Segal and Xiao (2011) and adapted to this context by Behrens et al. (2018b). The basis of my research forms a novel data set on German trade forecasts for the years 1970 to 2017 of four German economic research institutes, a collaboration of German forecasters and an international forecaster.

For most forecasters, I find evidence against the joint efficiency of trade forecasts. Only for one forecaster I can neither reject joint forecast efficiency for trade forecasts with a forecast horizon of half-a-year, nor for trade forecasts with a forecast horizon of one year. For two forecasters I find evidence against the joint efficiency of both half-a-year-ahead and one-year-ahead forecasts. The trade forecasts of the remaining forecasters are either jointly inefficient in the shorter or in the longer run. I compile a ranking of the six forecasters' joint efficiency of export and import forecasts. Finally, variable importance plots suggest that the forecasters use typical trade predictors more efficiently than other macroeconomic predictors, such as industrial orders, business climate surveys, consumer prices, and the unemployment rate. This leads to the assumption that a more efficient incorporation of these indicators could improve forecast accuracy, which could be pursued in future research. Improving forecast accuracy is an important task when it comes to policy making or investment decisions, as economic agents rely on professional forecasts when forming macroeconomic expectations (Carroll, 2003).

In future research, the approach of using nonparametric tree-based models to asses the properties of macroeconomic forecasts can be applied to forecasts for other variables and forecasts for other countries. It is also interesting to find out whether the underlying loss function of a trade forecast is of some other type than squared error loss, or to analyze possible behavioral biases in trade forecasts (see, for instance, Ito, 1990, on wishful expectations).

References

Andrade, P. / *Le Bihan,* H. (2013): “Inattentive professional forecasters”, Journal of Monetary Economics 60(8), 967–982.

Angelini, E. / *Bańbura,* M. / *Rünstler,* G. (2010): “Estimating and forecasting the Euro Area monthly national accounts from a dynamic factor model”, OECD Journal: Journal of Business Cycle Management and Analysis 1, 5–26.

Bae, K.-H. / *Stulz,* R. M. / *Tan,* H. (2008): “Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts”, Journal of Financial Economics 88(3), 581–606.

Banternghansa, C. / *McCracken,* M. W. (2009): “Forecast disagreement among FOMC members”, No. 2009-059A, Federal Reserve Bank of St. Louis Working Paper Series.

Behrens, C. / *Pierdzioch,* C. / *Risse,* M. (2018a): “A Bayesian analysis of the efficiency of growth and inflation forecasts for Germany”, SSRN Working Paper Series doi:dx.doi.org/10.2139/ssrn.3002394.

– (2018b): “A test of the joint efficiency of macroeconomic forecasts using multivariate random forests”, Journal of Forecasting 37(5), 560–572.

– (2018c): “Testing the optimality of inflation forecasts under flexible loss with random forests”, Economic Modelling 72, 270–277.

Berger, H. / *Ehrmann,* M. / *Fratzscher,* M. (2009): “Forecasting ECB monetary policy: Accuracy is a matter of geography”, European Economic Review 53(8), 1028–1041.

Breiman, L. (2001): “Random forests”, Machine learning 45(1), 5–32.

Carroll, C. D. (2003): “Macroeconomic expectations of households and professional forecasters”, The Quarterly Journal of Economics 118(1), 269–298.

D'Agostino, A. / *Modugno,* M. / *Osbat,* C. (2017): "A global trade model for the Euro Area", International Journal of Central Banking 13(4), 1–34.

Döhrn, R. / *Schmidt,* C. M. (2011): "Information or Institution? On the determinants of forecast accuracy", Journal of Economics and Statistics 231(1), 9–27.

Döpke, J. / *Fritsche,* U. (2006): "Growth and inflation forecasts for Germany - A panel-based assessment of accuracy and efficiency", Empirical Economics 31(3), 777–798.

Döpke, J. / *Fritsche,* U. / *Pierdzioch,* C. (2017): "Predicting recession with boosted regression trees", International Journal of Forecasting 33(4), 745–759.

Drechsel, K. / *Scheufele,* R. (2012): "The performance of short-term forecasts of the German economy before and during the 2008/2009 recession", International Journal of Forecasting 28(2), 428–445.

Elstner, S. / *Grimme,* C. / *Haskamp,* U. (2013): "Das ifo Exportklima - ein Frühindikator für die deutsche Exportprognose", ifo Schnelldienst 66(4), 36–43.

Engelke, C. / *Heinisch,* K. / *Schult,* C. (2019): "Effects of external assumptions on forecast errors", IWH Discussion Paper, forthcoming.

Frale, C. / *Marcellino,* M. / *Mazzi,* G. L. / *Proietti,* T. (2010): "Survey data as coincident or leading indicators", Journal of Forecasting 29(1-2), 109–131.

Frenkel, M. / *Lis,* E. M. / *Rülke,* J.-C. (2011): "Has the economic crisis of 2007-2009 changed the expectation formation process in the euro area?", Economic Modelling 28(4), 1808–1814.

Grimme, C. / *Lehmann,* R. / *Noeller,* M. (2018): "Forecasting imports with information from abroad", CESifo Group No. 7079.

Grossmann, S. H. / *Scheufele,* R. (2019): "PMIs: Reliable indicators for exports?", Review of International Economics, forthcoming.

Guichard, S. / *Rusticelli,* E. (2011): "A dynamic factor model for world trade growth", OECD Economics Department Working Papers No. 874.

Hacker, R. S. / *Hatemi-J,* A. (2004): "The effect of exchange rate changes on trade balances in the short and long run - Evidence from German trade with transitional Central European economies", Economics of Transition 12(4), 777–779.

Hastie, T. / *Tibshirani,* R. / *Friedman,* J. (2009): "The elements of statistical learning. Data mining, inference, and prediction", Springer: New York, 2.

Heilemann, U. / *Stekler,* H. O. (2013): "Has the accuracy of German macroeconomic forecasts improved?", German Economic Review 14(2), 235–253.

Heinisch, K. / *Scheufele,* R. (2018): "Bottom-up or direct? Forecasting German GDP in a data-rich environment", Empirical Economics 54, 705–745.

Hervé, K. / *Pain,* N. / *Richardson,* P. / *Sédillot,* F. / *Beffy,* P.-O. (2011): "The OECD's new global model", Economic Modelling 28(1-2), 289–601.

Hetemäki, l. / *Mikkola,* J. (2005): "Forecasting Germany's printing and writing paper imports", Forest Science 51(5), 483–497.

Holden, K. / *Peel,* D. A. (1990): "On testing for unbiasedness and efficiency of forecasts", The Manchester School 58(2), 120–127.

Ito, T. (1990): "Foreign exchange rate expectations: Micro survey data", The American Economic Review 80(3), 434–449.

Jakaitiene, A. / *Dées,* S. (2012): "Forecasting the world economy in the short term", The World Economy 35(3), 331–350.

James, G. / *Witten,* D. / *Hastie,* T. / *Tibshirani,* R. (2015): "An introduction to statistical learning", Springer: New York / Heidelberg.

Jannsen, N. / *Richter,* J. (2012): “Foreign capacity utilisation and German exports of capital goods”, Wirtschaftsdienst 92(12), 833–837.

Keck, A. / *Raubold,* A. / *Truppia,* A. (2009): “Forecasting international trade: A time series approach”, OECD Journal. Journal of Business Cycle Measurement and Analysis 2, 157–176.

Kirchgässner, G. (1984): “Wie gut sind die Prognosen der Arbeitsgemeinschaft wirtschaftswissenschaftlicher Forschungsinstitute in der Bundesrepublik Deutschland?”, Review of World Economics 120(2), 279–300.

– (1993): “Testing weak rationality of forecasts with different time horizons”, Journal of Forecasting 12(7), 541–558.

Kirchgässner, G. / *Müller,* U. K. (2006): “Are forecasters reluctant to revise their predictions? Some German evidence”, Journal of Forecasting 25(6), 401–423.

Kirchgässner, G. / *Savioz,* M. (2001): “Monetary policy forecasts for real GDP growth - An empirical investigation for the Federal Republic of Germany”, German Economic Review 2(4), 339–365.

Lehmann, R. (2015): “Survey-based indicators vs. hard data: What improves export forecasts in Europe?”, ifo Working Paper No. 196.

Malloy, C. J. (2005): “The geography of equity analysis”, Journal of Finance 60(2), 719–755.

McLachlan, G. J. (1999): “Mahalanobis distance”, Resonance 4(6), 20–26.

Mincer, J. A. / *Zarnowitz,* V. (1969): “The evaluation of economic forecasts”, in: Economic forecasts and expectations: Analysis of forecasting behavior and performance, ed. by Mincer, J. A., Economic forecasts and expectations: Analysis of forecasting behavior and performance, 3–46, National Bureau of Economic Research, New York.

Neumann, M. J. M. / *Buscher,* H. S. (1980): “Die Inflationsprognosen der Arbeitsgemeinschaft deutscher wirtschaftswissenschaftlicher Forschungsinstitute: Sind sie "rational"?”, Review of World Economics 116(3), 533–550.

Pesaran, H. M. / *Timmermann,* A. (1995): “Predictability of stock returns: Robustness and economic significance”, Journal of Finance 50(4), 1201–1228.

R Core Team (2018): R: A language and environment for statistical computing, R: version 3.5.1. R Foundation for Statistical Computing, Vienna, Austria, `https://www.R-project.org/`.

Rahman, R. / *Otridge,* J. / *Pal,* R. (2017): “IntegratedMRF: random forest-based framework for integrating prediction from different data types”, Bioinformatics 33(9), 1407–1410.

Riad, N. / *Errico,* L. / *Henn,* C. / *Saborowski,* C. / *Saito,* M. / *Turunen,* J. (2012): “Changing patterns of global trade”, International Monetary Fund, Washington D.C.

Segal, M. / *Xiao,* Y. (2011): “Multivariate random forests”, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1(1), 80–87.

Segal, M. R. (1992): “Tree-structured methods for longtitudinal data”, Journal of the American Statistical Association 87(418), 407 – 418.

Sinclair, T. / *Stekler,* H. O. / *Müller-Dröge,* H. C. (2016): “Evaluating forecasts of a vector of variables: A German forecasting competition”, Journal of Forecasting 35(6), 493–503.

Sinclair, T. M. / *Stekler,* H. O. (2013): “Examining the quality of early GDP component estimates”, International Journal of Forecasting 29(4), 736–750.

Sinclair, T. M. / *Stekler,* H. O. / *Carnow,* W. (2012): “A new approach for evaluating economic forecasts”, Economics Bulletin 32(3), 2332–2342.

– (2015): “Evaluating a vector of the Fed’s forecasts”, International Journal of Forecasting 31(1), 157–164.

Venables, W. N. / *Ripley,* B. D. (2002): “Modern applied statistics with S”, Springer, New York, fourth, `http://www.stats.ox.ac.uk/pub/MASS4`. ISBN 0-387-95457-0.

Vosen, S. / *Schmidt,* T. (2011): “Forecasting private consumption: Survey-based indicators vs. Google Trends”, Journal of Forecasting 30, 565–578.

Yu, L. / *Wang,* S. / *Lai,* K. K. (2008): "Forecasting China's foreign trade volume with a kernel-based hybrid econometric-AI ensemble learning approach", Journal of Systems and Complexity 21(1), 1–19.